A New Algorithm for Recursive Estimation of Parameters in Controlled ARMA Processes

Mayne, David Q.; Åström, Karl Johan

1981

CODEN: LUTFD2/(TFRT-3162)/1-040/(1981)

A NEW ALGORITHM FOR RECURSIVE ESTIMATION OF PARAMETERS IN CONTROLLED ARMA PROCESSES

K J ÅSTRÖM

D Q MAYNE

DEPARTMENT OF AUTOMATIC CONTROL
LUND INSTITUTE OF TECHNOLOGY
APRIL 1981

LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF AUTOMATIC CONTROL
Box 725
S 220 07 Lund 7 Sweden

Document name
Project report
Date of issue
April 1981
Document number
CODEN: LUTFD2/(TFRT-3162)/1-040/(1981)
Supervisor

Sponsoring organization

Author(s)
K J Åström
D Q Mayne

Title and subtitle

A new algorithm for recursive estimation of parameters in controlled ARMA processes

Abstract

A new recursive parameter estimation procedure for ARMA processes is proposed. The algorithm, which is based on convex approximations, is shown to be globally convergent.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
Enlgish
Number of pages
40
Security classification

Recipient's notes

Distribution: The report may be ordered from the Department of Automatic Control or borrowed through the University Library 2, Box 1010, S-221 03 Lund, Sweden, Telex: 33248 lubbis lund.

A NEW ALGORITHM FOR RECURSIVE ESTIMATION OF PARAMETERS IN CONTROLLED ARMA PROCESSES

K.J. Åström and D.Q. Mayne

1. INTRODUCTION

Mathematical models of processes in terms of controlled ARMA processes are of interest in control engineering. Such models are convenient representations of systems, whose input-output relation can be characterized by rational transfer functions, subject to disturbances having rational spectral densities. The problem of estimating the parameters of controlled ARMA processes has also received much attention. The maximum likelihood method was applied in [1] where it was shown that the maximum likelihood estimates were consistent, asymptotically efficient and asymptotically normal. It is a drawback of the maximum likelihood method that the likelihood function is nonlinear. This implies that there may be several local minima and that the optimization may be difficult. Various alternative methods for estimating the parameters in controlled ARMA processes have therefore been proposed, e.g. the generalized least squares [2], the extended least squares [3] and the two stage least squares [4]. A new method was proposed in [5] and [6]. This method is a multistep technique where least squares is used in each step. A recursive version of the method presented in [5] is presented and analysed in this report.

The recursive algorithm is of interest for the design of adaptive regulators and adaptive predictors. A review of recursive estimation methods is given in [7]. It may be legitimally questioned whether it is of any use to add yet another method to a large number of already existing routines. Thus question can be answered as follows. In the case of pure ARMA processes (i.e. no inputs) it is known that the maximum likelihood method is globally convergent. Many of the other recursive methods are, however, not globally convergent even for ARMA processes. There is, moreover, no method which is known to converge globally when inputs are present. One motivation for introducing the method presented in this report is that it is globally convergent. Another motivation is that the corresponding off-line method is consistent and asymptotically efficient.

It is unfortunately only possible to analyse the asymptotic properties of the recursive estimation procedure. Since short sample properties are also important it follows that it is not possible to evaluate estimation methods purely by analysis. For this reason it is thus necessary to explore the proposed method by simulation and also to investigate its numerical properties before it can be judged soberly.

The report is organized as follows. The process model, consisting of a controlled ARMA process, is described in Section 2. The off-line estimation method, which is the basis for the recursive algorithm, is presented in Section 3. The properties of the off-line method are also discussed briefly in that section. The recursive algorithm is presented in Section 4, and its properties are analysed in Section 5. It is shown that the algorithm will under reasonable assumptions always converge.

2. PROCESS DESCRIPTION

The process is described in this section. The assumptions made on the process are also stated.

Only discrete time processes are discussed. It is thus assumed that time T is the set of integers $T = \{\ldots,-1,0,1,\ldots\}$. Signals are functions from T to R. They are denoted by lower case letters like u and y. It is assumed that the process is a controlled ARMA process described by:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \tag{2.1}$$

where u is the input signal, y the output signal and $\{e(t)\}$ is a sequence of independent identically distributed random variables with zero mean values and covariances λ^2. In (2.1) $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are polynomials of degree n in the backward shift operator q^{-1}:

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n}$$

$$B(q^{-1}) = b_1q^{-1} + \dots + b_nq^{-n}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \dots + c_nq^{-n} \tag{2.2}$$

It is assumed that there are no factors common to A, B and C. There is no loss of generality is assuming that $A(0) = C(0) = 1$. The assumption that $B(0) = 0$ is not important; it is made only to obtain symmetry in certain equations.

The model (2.1) will also be written in the following abbreviated form:

$$Ay = Bu + Ce \tag{2.1'}$$

The problem we will consider is estimation of the parameters of the process (2.1) from observations of input-output records $\{u(t), y(t), t \in T\}$ of the process. For easy reference the parameters are gathered into the vector:

$$\Theta = \text{coeff}\,[A(q^{-1}), B(q^{-1}), C(q^{-1})]$$

$$\triangleq [a_1\ a_2\ \ldots\ a_n\ b_1\ b_2\ \ldots\ b_n\ c_1\ c_2\ \ldots\ c_n] \tag{2.3}$$

Conversely to each parameter vector Θ we can also associate polynomials A, B and C given by:

$$A(q^{-1}) = 1 + \Theta_1 q^{-1} + \ldots + \Theta_n q^{-n}$$

$$B(q^{-1}) = \Theta_{n+1} q^{-1} + \ldots + \Theta_{2n} q^{-n} \tag{2.4}$$

$$C(q^{-1}) = 1 + \Theta_{2n+1} q^{-1} + \ldots + \Theta_{3n} q^{-n} \quad .$$

Assumptions

The following assumptions are made on the process (2.1) which generates the data to be used in the system identification:

A1: There are no factors common to all of A, B and C and $C(q^{-1}) \neq 1$. Furthermore the zeros of the polynomials $z^n A(z^{-1})$ and $z^n C(z^{-1})$ are strictly less than $\gamma < 1$ in magnitude.

A2: The input signal is covariance stationary and persistently exciting of arbitrary order p.

A3: The disturbance e is a sequence of independent random variables with bounded fourth moments. The disturbance is uncorrelated.

3. THE OFF-LINE ALGORITHM

A new algorithm for estimating the parameters of the process (2.1) was proposed in [5] and [6]. The algorithm is a multistep method where least squares is used in each step. The algorithm is described in this section. Some of its properties are also discussed together with minor modifications which lead to different versions of the algorithm.

The Algorithm

Step 1: Estimate A_1 and B_1 in the model

$$A_1 y = B_1 u \tag{M_1}$$

where $\deg A_1 = \deg B_1 = p$ by least squares. Let ε_1 denote the residuals obtained.

Step 2: Estimate A_2, B_2 and C_2 in the model

$$A_2 y = B_2 u + C_2 \varepsilon_1 \tag{M_2}$$

where the polynomials have the form (2.2) and ε_1 is the residual from Step 1, by least squares.

Step 3: Filter the signals u, y and ε_1 by C_2^{-1} where C_2 is the polynomial estimated in Step 2 to obtain

$$\bar{y} = \frac{1}{C_2} y, \quad \bar{u} = \frac{1}{C_2} u, \quad \bar{\varepsilon}_1 = \frac{1}{C_2} \varepsilon_1$$

Then estimate A_3, B_3 and C_3 in the model

$$A_3\bar{y} = B_3\bar{u} + C_3\bar{\varepsilon}_1 \qquad (\mathcal{M}_3)$$

where the polynomials A_3, B_3 and C_3 are of the form (2.2), by least squares.

Minor Modifications

In Step 1 the parameters are determined in such a way that the criterion

$$\sum_t [A_1(q^{-1})y(t) - B_1(q^{-1})u(t)]^2$$

is minimized. Similarly the criterion

$$\sum_t [A_2(q^{-1})y(t) - B_2(q^{-1})u(t) - C_2(q^{-1})\varepsilon_1(t)]^2$$

is used in Step 2. Since the leading coefficient in $C_2(q^{-1})$ is one (c.f. (2.2)) the alternative criterion

$$\sum_t \{A_2(q^{-1})y(t) - B_2(q^{-1})u(t) - [C_2(q^{-1}) - 1]\varepsilon_1(t)\}^2$$

can also be used in Step 2. This means that the model $\mathcal{M}_2$ is replaced by

$$A_2 y = B_2 u + (C_2 - 1)\varepsilon_1 \qquad (\mathcal{M}_2')$$

Similarly the model $\mathcal{M}_3$ in Step 3 can be replaced by

$$A_3\bar{y} = B_3\bar{u} + (C_3 - C_2)\bar{\varepsilon}_1 \qquad (\mathcal{M}_3')$$

It is easily seen that these modifications are asymptotically (for large p and large N) unimportant. They will, however, give versions of the algorithm that are slightly different for finite p and N.

Properties of the Off-Line Algorithm

The residuals ε_1 obtained in Step 1 will be close to the process innovations e if p and N are large. In [5] and [6] it is shown that if all zeros of the polynomial $z^n C(z^{-1})$ have magnitudes less than $\gamma < 1$ then asymptotically for large N

$$E|\varepsilon_1(t)-e(t)|^2 \leq K\,\gamma^{2(p-n)}$$

For large N and p the polynomials A_1 and B_1 obtained in Step 1 are also close to the polynomials obtained by truncating the series expansions in q^{-1} of the rational functions $A(q^{-1})/C(q^{-1})$ and $B(q^{-1})/C(q^{-1})$ in a sense given precisely in [6]. This result is further illustrated by the following simple example:

Example 3.1

Consider the process

$$y(t) = e(t) + ce(t-1), \quad |c| < 1$$

A straightforward solution of the normal equations gives asymptotically, for large N, the following coefficients of the polynomial A_1

$$a_i = (-c)^i \, \frac{1-c^{2(p-i+1)}}{1-c^{2(p+1)}} \quad , \; i = 1, \ldots, p$$

The truncation of the rational function $1/C(q^{-1})$ gives a polynomial with the coefficients $(-c)^i$, $i = 1, \ldots, p$. □

An estimate of the parameters of the process (2.1) is already obtained in Step 2. This estimate is the well known two-stage least squares (2SLS) estimate which was originally proposed by Durbin [4], and later in the automatic control literature [8] and [9]. (Note that the term 2LSL is used in the econometrics literature [10] to describe a different estimate, closely related to instrumental variable estimators.) It is shown in [6] and [9] that the bias of the 2SLS estimate can be made arbitrarily small by choosing the parameter p in the first step sufficiently large. The following example from [9] gives the bias of the 2SLS estimate of a first order system:

Example 3.2

Consider the process

$$y(t) + ay(t-1) = bu(t-1) + e(t) + ce(t-1)$$

where the input u is white noise which is independent of e and has variance σ^2. Then [9] as $N \to \infty$, the 2SLS estimate converges to

$$\hat{a} = a - \frac{\delta c^{2p-1}(1-c^2)}{(1-c^{2p})(1+\delta-c^{2p})}$$

$$\hat{b} = b$$

$$\hat{c} = c - \frac{c^{2p-1}(1-c^2)}{(1-c^{2p})}$$

$$\delta = \frac{\lambda^2(1-a^2)(1-c^2)}{\lambda^2(c-a)^2+b^2\sigma^2}$$

If the modified estimate obtained by fitting the model M_2' instead of M_2 then [9] the 2SLS estimate converges instead to

$$\hat{a} = a$$

$$\hat{b} = b$$

$$\hat{c} = c - \frac{c^{2p+1}(1-c^2)}{1+\delta-c^{2p+2}}$$

□

The example thus indicates that it is slightly advantageous (at least for n=1) to use the version of the algorithm where the model M_2' is fitted in the second step.

The 2SLS estimate is known to be inefficient. The purpose of the third step in the algorithm is to reduce the variance of 2SLS estimate.

In the second stage of the algorithm the loss function

$$V_2(\theta_2) = \frac{1}{2N}\sum_{t=1}^{N} [A_2 y(t) - B_2 u(t) - C_2\varepsilon_1(t)]^2$$

is minimised with respect to θ_2, i.e. with respect to (the coefficients of) A_2, B_2, C_2. In the third stage the loss function

$$V_3(\theta_3) = \frac{1}{2N}\sum_{t=1}^{N} [\frac{A_3}{C_2}\ y(t) - \frac{B_3}{C_2}\ u(t) - \frac{C_3}{C_2}\ \varepsilon_1(t)]^2$$

is minimised with respect to θ_3 (i.e. A_3, B_3, C_3). The likelihood function associated with the problem is:

$$V_{ML}(\theta) = \frac{1}{2N}\sum_{t=1}^{N} [\frac{A}{C}\, y(t) - \frac{B}{C}\, u(t)]^2$$

The maximum likelihood estimate θ_{ML} minimises $V_{ML}(\theta)$ with respect to θ. It is shown in [14] that $\theta_3 \to \theta^p$ w.p. 1, as $N \to \infty$, that $\theta^p \to \theta$ as $p \to \infty$, that the asymptotic variance (as $N \to \infty$) of $\sqrt{N}\,[\theta_3 - \theta^p]$ is:

$$[E\nabla^2 V_3(\theta^p)]^{-1} \quad [EN\nabla V_3(\theta^p) \ \nabla V_3^T(\theta^p)]^{-1} \ [E\nabla^2 V_3(\theta^p)]^{-1}$$

and that $E\nabla^2 V_3(\theta^p) \to E\nabla^2 V_{ML}(\theta)$ and $EN\nabla V_3(\theta^p) \ \nabla V_3^T \ (\theta^p) \to E\nabla^2 V_{ML}(\theta)$ as $p \to \infty$. Hence the asymptotic variance of $\sqrt{N}\,[\theta_3 - \theta^p]$ tends to that of $\sqrt{N}\,[\theta_{ML} - \theta]$ as $p \to \infty$; this property is not possessed by θ_2. Since V_3 is a convex (indeed quadratic) function, the estimation scheme can be regarded as based on a convex approximation of the likelihood function.

4. THE RECURSIVE ALGORITHM

Since the algorithm given in the previous section is composed of three least squares steps it is easy to obtain a recursive algorithm simply be replacing the least square estimations by recursive least squares in each step. The recursive algorithm will then be built up of three recursive least squares steps. The equations will be very similar for all steps. The following notation is introduced to describe the algorithm.

$$\theta_1 = \text{coeff } (A_1, B_1), \quad \dim \theta_1 = 2p$$

$$\theta_2 = \text{coeff } (A_2, B_2, C_2), \quad \dim \theta_2 = 3n$$

$$\theta_3 = \text{coeff } (A_3, B_3, C_3), \quad \dim \theta_3 = 3n \tag{4.1}$$

The regressors are denoted as

$$z_1(t) = [-y(t-1)\ldots -y(t-p) \ \ u(t-1)\ldots u(t-p)]^T$$

$$z_2(t) = [-y(t-1)\ldots -y(t-n) \ \ u(t-1)\ldots u(t-n) \ \ \varepsilon_1(t-1)\ldots \varepsilon_1(t-n)]^T$$